semana 10

Abigail Ramos

13/10/2022

```
## Loading required package: splines
## Loading required package: RcmdrMisc
## Loading required package: car
## Loading required package: carData
## Loading required package: sandwich
## Loading required package: effects
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
## La interfaz R-Commander sólo funciona en sesiones interactivas
##
## Attaching package: 'Rcmdr'
## The following object is masked from 'package:base':
##
##     errorCondition
```

Intervalo de confianza para la media en una población normal con varianza conocida

Supuesto Práctico 1

El archivo empleados.xls nos informa de la edad, altura, peso, sexo y posesión de coche de 100 empleados de una empresa. Suponiendo la normalidad de la variable Altura, calcular el intervalo de confianza sobre la altura media poblacional a un 95% de confianza, sabiendo que la varianza poblacional es 6.

Importar datos de Excel a R-Commander

```
> empleados <-
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/empleados.xls"
+   rownames=FALSE, header=TRUE, na="", sheet="Respuestas", stringsAsFactors=TRUE)
```

A continuación guardamos el archivo (Datos/Conjunto de datos activo/Guardar el conjunto de datos activo) como fichero Empleados.RData.

```
> save("empleados",
+   file="C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/empleados.R.Data
```

Introducimos en R los datos, que proporciona el enunciado, relativos al nivel de significación y la varianza poblacional de la variable.

```
> alpha<-0.05
> varianza<-7.5
```

Calculamos por separado cada uno de los elementos restantes que necesitamos para obtener el intervalo de confianza.

```
> n <- nrow(empleados)
> media <- mean(empleados$Altura)
>
> cuantil<- qnorm(1-alpha/2)
```

Por último, calculamos los extremos inferior y superior del intervalo de acuerdo

```
> lim_inferior<-media -cuantil * sqrt(varianza) / sqrt(n)
> lim_inferior
```

```
[1] 176.4605
```

```
> lim_superior<- media + cuantil * sqrt(varianza) / sqrt(n)
> lim_inferior
```

```
[1] 176.4605
```

Por lo que el intervalo de confianza que buscamos es (176.4605, 177.5395).

Intervalo de confianza para la media en una población normal con varianza desconocida

Supuesto Práctico 2

Considerando el conjunto de datos de empleados.xls y asumiendo que la variable que mide la altura de los empleados sigue una distribución Normal con varianza desconocida. Calcular un intervalo de confianza a un nivel de confianza del 90% para la altura media poblacional.

Accedemos al menú Test t para una muestra, seleccionando en el menú principal: Estadísticos/ Medias/ Test t para una muestra.

```
> with(empleados, (t.test(Altura, alternative='two.sided', mu=0.0, conf.level=.90)))
```

```
    One Sample t-test

data:  Altura
t = 214.41, df = 98, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 175.6292 178.3708
sample estimates:
mean of x
      177
```

En la parte superior izquierda se muestra una lista con todas las variables cuantitativas del archivo de datos que son susceptibles de ser contrastadas, de la cual debemos elegir exclusivamente una.

El intervalo de confianza pedido es (175.6292, 178.3708).

Intervalo de confianza para la proporción

Supuesto Práctico 3

A partir del conjunto de datos de empleados.xls, obtener un intervalo de confianza al 95% para la proporción de empleados varones en la población.

Solución

Accedemos al menú Test de proporciones para una muestra de R-Commander, seleccionando en el menú principal: Estadísticos/Proporciones/ Test de proporciones para una muestra.

En la pestaña Datos del cuadro de diálogo, se muestra una lista con todas las variables cualitativas que pueden utilizarse en este tipo de contrastes, de entre las cuales tenemos que elegir una. En este caso elegimos Sexo.

```
> local({
+   .Table <- xtabs(~ Sexo , data= empleados )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.95, correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
Sexo
Hombre  Mujer
    87     12

	1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 56.818, df = 1, p-value = 4.78e-14
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.7999934 0.9292845
sample estimates:
        p
0.8787879
```

En este ejemplo, las dos posibles opciones para la variable Sexo son "Hombre" y "Mujer", por lo que la primera de las categorías para R-Commander es "Hombre". Dado que la hipótesis que se ha planteado se ha hecho sobre los hombres no es necesario hacer ninguna modificación.

Si, por el contrario, la hipótesis del problema se hubiera planteado sobre las mujeres, deberíamos hacer una recodificación previa de la variable para situar la categoría "Mujer" como la primera.

En la segunda parte se muestran los resultados del contraste de hipótesis, que analizaremos en la práctica sobre Contrastes de Hipótesis.

Por lo que el intervalo de confianza, a un nivel de confianza del 95% para la proporción de empleados varones en la población es (0.7999934, 0.9292845).

Intervalo de confianza para el cociente de varianzas en dos poblaciones normales independientes

Supuesto Práctico 4

Continuando con los datos del archivo empleados.xls y asumiendo que el peso en hombres y el peso en mujeres se distribuyen según distribuciones normales con medias y varianzas desconocidas. Calcular un intervalo de confianza a un nivel de confianza del 95% para el cociente de varianzas en ambas poblaciones. ¿Puede asumirse que ambas varianzas son iguales?

Accedemos a Test F para dos varianzas, de R-Commander, seleccionando en el menú principal: Estadísticos/ Varianzas/ Test F para dos varianzas

La pestaña Datos muestra dos listas de variables.

- La lista de Grupos, muestra todas las variables cualitativas del fichero de datos. En esta lista tenemos que seleccionar cuál es la variable que nos va a dividir la muestra de observaciones en dos submuestras independientes. En nuestro caso, al tratar con hombres y mujeres, seleccionamos el Sexo del empleado como variable de agrupación.
- La lista de Variable explicativa, muestra todas las variables cuantitativas del fichero de datos. Señalamos la variable principal sobre la cual se va a llevar a cabo el contraste (en nuestro caso, Peso).

```
> Tapply(Peso ~ Sexo, var, na.action=na.omit, data=empleados) # variances by group
```

```
   Hombre     Mujer 
166.99626  92.06061 
```

```
> var.test(Peso ~ Sexo, alternative='two.sided', conf.level=.95, data=empleados)
```

```

    F test to compare two variances

data:  Peso by Sexo
F = 1.814, num df = 86, denom df = 11, p-value = 0.2752
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6112226 3.8937784
sample estimates:
ratio of variances 
          1.813982 
```

Esta salida, muestra que el intervalo de confianza para el cociente de las varianzas es (0.6112226, 3.8937784).

La interpretación del intervalo de confianza puede servirnos para concluir acerca de la igualdad de las varianzas. En este ejemplo, dicho intervalo es (0.6112226, 3.8937784) que, como podemos comprobar incluye al 1 entre sus posibles valores. Esto implica que a un nivel de confianza del 95% se puede suponer que el cociente entre las dos varianzas puede tomar el valor 1 o, lo que es lo mismo, que las dos varianzas son iguales.

Una vez se ha determinado la igualdad (o desigualdad) de las varianzas de ambas distribuciones, procedemos a calcular el intervalo de confianza para la diferencia de las medias propiamente dicho.

a) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas pero supuestas iguales

b) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas, distintas y tamaños muestrales grandes

Supuesto Práctico 5

Sabiendo que las varianzas son iguales (Supuesto práctico 4), obtener un intervalo de confianza al 95% para la diferencia del peso medio entre hombres y mujeres. ¿Puede suponerse que el peso medio entre hombres y mujeres es igual?

Para obtener el Test t para muestras independientes con R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para muestras independientes.

La pestaña Datos muestra dos listas de variables. Como ya se ha comentado con anterioridad, en la lista de la izquierda (Grupos) tenemos que escoger la variable a partir de la cual se formarán los dos grupos de observaciones, seleccionamos Sexo. En la de la derecha (Variable explicada) seleccionamos la variable cuya diferencia de medias en las poblaciones queremos estudiar, seleccionamos Peso.

```
> t.test(Peso~Sexo, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=empleados)
```

```
    Welch Two Sample t-test

data:  Peso by Sexo
t = 3.8302, df = 17.057, p-value = 0.001333
alternative hypothesis: true difference in means between group Hombre and group Mujer is not equal to 0
95 percent confidence interval:
  5.329686 18.394452
sample estimates:
mean in group Hombre  mean in group Mujer
            76.19540             64.33333
```

Se puede afirmar que el intervalo de confianza a un 95% de confianza para la diferencia de las medias del peso medio para hombres y mujeres es (4.167581, 19.556557). Como el 0 no está dentro de este intervalo, no tenemos suficiente evidencia muestral para decir que el peso medio de hombres y mujeres sea el mismo.

Intervalo de confianza para la diferencia de medias en dos poblaciones normales relacionadas

Supuesto Práctico 6

Se desea evaluar la eficacia de un fármaco para la reducción del nivel de glucosa en pacientes. Para ello, se selecciona una muestra de 10 pacientes a los que se les mide su nivel de glucosa en sangre antes y después del suministro del medicamento. Los resultados aparecen recogidos en la siguiente tabla:

Solución

En primer lugar, debemos crear un nuevo conjunto de datos con la información que nos proporciona la tabla. El conjunto de datos estará formado por dos variables con los niveles de glucosa antes y después de la aplicación del fármaco. Organizamos los datos en un fichero .txt (supuesto6.txt)

A continuación seleccionamos en el menú principal: Datos/Importar datos/desde archivo de texto

```
> supuesto6 <-
+   read.table("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/supuesto6.t
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
```

A continuación, accedemos al menú Test t para datos relacionados en R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para datos relacionados

Es importante destacar que, a diferencia del caso de muestras independientes, cuando trabajamos con muestras pareadas no necesitamos una variable de agrupación, sino que debemos seleccionar las dos variables a analizar de forma separada.

En la pestaña Opciones personalizamos el contraste conforme al problema que estemos resolviendo. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado.

```
> with(supuesto6, (t.test(Antes, Después, alternative='two.sided', conf.level=.90,
+   paired=TRUE)))
```

```
    Paired t-test

data:  Antes and Después
t = -2.1581, df = 5, p-value = 0.08338
alternative hypothesis: true mean difference is not equal to 0
90 percent confidence interval:
```

```
 -2.9972387 -0.1027613
sample estimates:
mean difference
          -1.55
```

En la segunda parte de los resultados se incluye el intervalo de confianza al 90% para la diferencia de las medias de la variable glucosa, que es (-9.290501, 16.790501). Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambos niveles medios de glucosa son iguales.

Intervalo de confianza para la diferencia de dos proporciones

Supuesto Práctico 7

A partir del conjunto de datos empleados.xls, obtener un intervalo de confianza al 85% para la diferencia entre la proporción de empleados hombres y mujeres que tienen coche. ¿Pueden considerarse ambas proporciones iguales?

Solución

Dado que la hipótesis se ha planteado sobre la proporción de hombres y mujeres que tienen coche, es necesario recodificar la variable coche. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

En esta pantalla:

- Se selecciona Coche en Variables a recodificar
- En Nuevo nombre o prefijo para variables múltiples recodificadas hemos puesto como nombre de la variable: coche_rec
- En la ventana Introducir directrices de recodificación: "No"="2No" y "Sí"="1Sí"

```
> empleados <- within(empleados, {
+   coche_rec <- Recode(Coche, '"No"="2No"; "Sí"="1Sí";', as.factor=TRUE)
+ })
```

Accedemos al menú Test de proporciones para dos muestras de R-Commander, seleccionando en el menú principal: Estadísticos/Proporciones/ Test de proporciones para dos muestras

La pestaña Datos muestra dos listas con las variables cualitativas que incluye el conjunto de datos. De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Sexo, ya que distinguimos entre hombres y mujeres) y de la segunda, la variable de interés (que es, si el empleado tiene coche, coche_rec).

En la pestaña Opciones indicamos el nivel de confianza propuesto en el enunciado, 85%

```
> library(abind, pos=17)
```

```
> local({  .Table <- xtabs(~Sexo+coche_rec, data=empleados)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.85, correct=FALSE)
+ })
```

```
Percentage table:
        coche_rec
Sexo      1Sí  2No Total Count
  Hombre 52.9 47.1   100    87
  Mujer  50.0 50.0   100    12
```

```
    2-sample test for equality of proportions without continuity correction

data:  .Table
X-squared = 0.03492, df = 1, p-value = 0.8518
alternative hypothesis: two.sided
85 percent confidence interval:
 -0.1928653  0.2503366
sample estimates:
   prop 1    prop 2
0.5287356 0.5000000
```

El intervalo de confianza pedido al 85% de confíanza es (-0.2729805, 0.3304517). El 0 está dentro de este intervalo, por lo que podemos concluir que las proporciones de hombre y mujeres que tienen coche coinciden.

EJERCICIOS GUIADOS

Ejercicio Guiado 1

En la base de datos universidad.txt tenemos información sobre las variables, Coeficiente intelectual y Asistir a clase de estadística de dos grupos de alumnos, dependiendo del turno de clase en el que se encuentren. El turno de mañana se define como A y el de tarde como B.

Se pide:

a) Obtener un intervalo de confianza a un nivel del 99% para el cociente intelectual medio, sabiendo que la varianza poblacional es igual a 3

b) Obtener un intervalo de confianza a un nivel del 95% para el cociente intelectual medio

c) Obtener un intervalo de confianza a un nivel del 98% para la diferencia media de cociente intelectual entre el grupo A y B. ¿Puede suponerse que el cociente intelectual medio entre ambos grupos es igual?

d) Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo A. Y un intervalo de confianza al 90% para la proporción de alumnos en el grupo B

e) Obtener un intervalo de confianza a un nivel del 93% para la diferencia entre la proporción de alumnos en el grupo A y B que no tienen clase de estadística.

Solución

```
> universidad <-
+   read.table("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/universidad
+   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
```

a) Obtener un intervalo de confianza a un nivel del 99% para el cociente intelectual medio, sabiendo que la varianza poblacional es igual a 3

En este caso nos encontramos ante un intervalo de confianza sobre la media de una población normal con varianza conocida, por lo que hay que calcularlo mediante código

Introducimos en R Commander los datos relativos al nivel de significación y la varianza poblacional de la variable que proporciona el enunciado.

```
> alpha<-0.01
> varianza<-3
```

Calculamos por separado cada uno de los elementos restantes que necesitamos para obtener el intervalo de confianza.

```
> n <- nrow(universidad)
> media <- mean(universidad$C.I.)
> cuantil<- qnorm(1 - alpha/2)
```

Por último, calculamos los extremos inferior y superior del intervalo

```
> lim_inferior<-media - cuantil * sqrt(varianza) / sqrt(n)
> lim_inferior
```

```
[1] 99.20477
```

```
> lim_superior<- media + cuantil * sqrt(varianza) / sqrt(n)
> lim_superior
```

```
[1] 103.1952
```

Por lo que el intervalo de confianza que buscamos es **(99.20477, 103.1952).**

Nota: Recordar que para que se ejecute una instrucción hay que pulsar Ejecutar

b) Obtener un intervalo de confianza a un nivel del 95% para el cociente intelectual medio

En este caso nos encontramos ante un intervalo de confianza sobre la media de una población normal con varianza desconocida. Para realizarlo con R-Commander, seleccionamos en el menú principal: Estadísticos/Medias/ Test t para una muestra

En la parte superior izquierda aparece una lista con todas las variables cuantitativas del archivo de datos que son susceptibles de ser contrastadas, de la cual debemos elegir exclusivamente una. Elegimos C.I

El resto de opciones las dejamos por defecto, ya que el enunciado nos pide el intervalo de confianza al 95%.

```
> with(universidad, (t.test(C.I., alternative='two.sided', mu=0.0, conf.level=.95)))
```

```
    One Sample t-test

data:  C.I.
t = 79.024, df = 4, p-value = 1.537e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  97.64442 104.75558
sample estimates:
mean of x
    101.2
```

Por lo que el intervalo de confianza es **(97.64442, 104.75558).**

c) Obtener un intervalo de confianza a un nivel del 98% para la diferencia media de cociente intelectual entre el grupo A y B. ¿Puede suponerse que el cociente intelectual medio entre ambos grupos es igual?

En este caso nos encontramos ante un intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes.

En primer lugar tenemos que saber si las varianzas de ambas distribuciones son iguales. Para ello, seleccionamos en R-Commander: Estadísticos/ Varianzas/ Test F para dos varianzas. Se muestra la siguiente pantalla

La pestaña Datos muestra dos listas de variables. La lista de la izquierda (Grupos) incluye todas las variables cualitativas del fichero de datos. En esta lista tenemos que seleccionar cuál es la variable que nos va a dividir la muestra de observaciones en dos submuestras independientes. En nuestro caso, Grupo. En la lista de la derecha se incluyen las variables cuantitativas del fichero de datos. Aquí tenemos que señalar la variable principal sobre la cual se va a llevar a cabo el contraste (en nuestro caso, C.I.). Seleccionamos Opciones

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, sólo hemos de mirar el nivel de confianza. Por defecto es 0.95, por lo que lo cambiamos a 0.98. Pulsamos Aceptar y se muestra la siguiente salida.

```
> Tapply(C.I. ~ Grupo, var, na.action=na.omit, data=universidad) # variances by group
```

```
       A        B 
12.33333  8.00000 
```

```
> var.test(C.I. ~ Grupo, alternative='two.sided', conf.level=.98, data=universidad)
```

```
    F test to compare two variances

data:  C.I. by Grupo
F = 1.5417, num df = 2, denom df = 1, p-value = 0.9897
alternative hypothesis: true ratio of variances is not equal to 1
98 percent confidence interval:
 3.083642e-04 1.518580e+02
sample estimates:
ratio of variances 
          1.541667 
```

El intervalo de confianza para el cociente de las varianzas, **(3.083642e-04, 1.518580e+02)**. Dicho intervalo incluye al 1 entre sus posibles valores. Esto implica que a un nivel de confianza del 98% se puede suponer que el cociente entre las dos varianzas puede tomar el valor 1 o, lo que es lo mismo, que las dos varianzas son iguales.

Una vez se ha determinado la igualdad de las varianzas de ambas distribuciones, procedemos a calcular el intervalo de confianza para la diferencia de las medias cuando las varianzas poblacionales son iguales. Para ello, seleccionamos en R-Commander: Estadísticos/Medias/Test t para muestras independientes

La pestaña Opciones muestra todas las opciones del contraste que podemos modificar. En nuestro caso como sólo buscamos el intervalo de confianza, especificamos el nivel de confianza que se va a asumir al calcular el intervalo (98%) e indicamos si las varianzas de las dos poblaciones pueden suponerse iguales o no, en nuestro caso Sí.

```
> t.test(C.I.~Grupo, alternative='two.sided', conf.level=.98, var.equal=TRUE, data=universidad)
```

```
    Two Sample t-test

data:  C.I. by Grupo
t = 0.11066, df = 3, p-value = 0.9189
alternative hypothesis: true difference in means between group A and group B is not equal to 0
98 percent confidence interval:
 -13.34472  14.01139
sample estimates:
mean in group A mean in group B 
       101.3333        101.0000 
```

El intervalo de confianza a un 98% de confianza para la diferencia de las medias del cociente intelectual entre el grupo A y el B es **(-13.34472, 14.01139)**. Como el 0 está dentro de este intervalo, tenemos suficiente evidencia muestral para decir que el cociente intelectual medio del grupo A y B son iguales.

d) Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo A. Y un intervalo de confianza al 90% para la proporción de alumnos en el grupo B

En este caso nos encontramos ante un intervalo de confianza para la proporción.

- Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo A.

Dado que la hipótesis que se ha planteado se ha hecho sobre el grupo A, no es necesario hacer ninguna recodificación de la variable. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para una muestra

En este caso elegimos la variable Grupo. En la pestaña Opciones modificamos el valor del nivel de confianza a un 90%.

```
> local({
+   .Table <- xtabs(~ Grupo , data= universidad )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.90, correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
Grupo
A B
3 2

Warning in prop.test(rbind(.Table), alternative = "two.sided", p = 0.5, : Chi-
squared approximation may be incorrect

    1-sample proportions test without continuity correction


data:  rbind(.Table), null probability 0.5
X-squared = 0.2, df = 1, p-value = 0.6547
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.2724832 0.8572935
sample estimates:
  p
0.6
```

Por lo que el intervalo de confianza, a un nivel de confianza del 90% para la proporción de estudiantes en el grupo A es **(0.2724832, 0.8572935).**

- Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo B.

Dado que la hipótesis que se ha planteado se ha hecho sobre el grupo B es necesario hacer una recodificación de la variable. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

```
> universidad <- within(universidad, {
+   Grupo_rec <- Recode(Grupo, '"A"="2A"; "B"="1B"', as.factor=TRUE)
+ })
```

Una vez recodificada la variable, pasamos a calcular el intervalo de confianza. En este caso nos encontramos ante un intervalo de confianza para la proporción. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para una muestra.

En este caso elegimos la variable Grupo_rec. En la pestaña Opciones modificamos el valor del nivel de confianza a un 90%.

```
> local({
+   .Table <- xtabs(~ Grupo_rec , data= universidad )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.90, correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
```

```
Grupo_rec
1B 2A
 2  3

Warning in prop.test(rbind(.Table), alternative = "two.sided", p = 0.5, : Chi-
squared approximation may be incorrect

	1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 0.2, df = 1, p-value = 0.6547
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.1427065 0.7275168
sample estimates:
  p
0.4
```

Por lo que el intervalo de confianza, a un nivel de confianza del 90% para la proporción de estudiantes en el grupo B es **(0.1427065, 0.7275168).**

e) Obtener un intervalo de confianza a un nivel del 93% para la diferencia entre la proporción de alumnos en el grupo A y B que no tienen clase de estadística.

Dado que la hipótesis que se ha planteado se ha hecho sobre los alumnos que no tienen clase de estadística no es necesario hacer una recodificación de la variable Estadística. En este caso nos encontramos ante un intervalo de confianza para la diferencia de dos proporciones. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para dos muestras.

De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Grupo) y de la segunda, la variable de interés (Estadística). Ya en la segunda pestaña (Opciones) indicamos el nivel de confianza propuesto en el enunciado, 93%.

```
> local({  .Table <- xtabs(~Grupo+Estadistica, data=universidad)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.93, correct=FALSE)
+ })
```

```
Percentage table:
     Estadistica
Grupo   No S\xed Total Count
    A 33.3  66.7   100     3
    B 50.0  50.0   100     2

Warning in prop.test(.Table, alternative = "two.sided", conf.level = 0.93, :
Chi-squared approximation may be incorrect

	2-sample test for equality of proportions without continuity correction

data:  .Table
X-squared = 0.13889, df = 1, p-value = 0.7094
alternative hypothesis: two.sided
93 percent confidence interval:
 -0.9750999  0.6417665
sample estimates:
   prop 1    prop 2
0.3333333 0.5000000
```

Por último, el programa devuelve el intervalo de confianza al 93% de confíanza, **(-0.9750999, 0.6417665)**. El 0 está dentro de este intervalo, por lo que podemos concluir que las proporciones de alumnos en ambos grupos que no tienen clase de estadística coinciden.

Ejercicio Guiado 2* (Resuelto)

En un hospital se elige una muestra de pacientes y se les mide la tasa cardíaca por la mañana (TCM) y a última hora de la tard (TCT).

Estudiar mediante un intervalo de confianza al 99% si, por término medio, la tasa cardíaca es igual por la mañana y a última hora de la tarde.

Solución

```
> tcardiaca <-
+   read.table("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/tcardiaca.t
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
```

En este caso nos encontramos ante un intervalo de confianza para la diferencia medias en dos poblaciones normales relacionadas. Seleccionamos: Estadísticos/Medias/Test t para datos relacionados.

La pestaña Datos muestra dos listas de variables cada una de las cuales incluye todas las variables cuantitativas que son susceptibles de ser analizadas. Seleccionamos en cada lista la variable que nos interese (TCM y TCT en nuestro caso).

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado (99%).

```
> with(tcardiaca, (t.test(TCM, TCT, alternative='two.sided', conf.level=.99, paired=TRUE)))
```

```
    Paired t-test

data:  TCM and TCT
t = -1.6236, df = 4, p-value = 0.1798
alternative hypothesis: true mean difference is not equal to 0
99 percent confidence interval:
 -18.411312   8.811312
sample estimates:
mean difference
           -4.8
```

En la segunda parte de los resultados se incluye el intervalo de confianzaal 99% para la diferencia de las medias de la variable tasa cardíaca, que es **(-18.411312, 8.811312)**. Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambas tasas cardíacas son iguales.

Ejercicio Guiado 3* (Resuelto)

Una determinada empresa quiere saber si su nuevo producto tendrá más aceptación en la población de hombres o entre las mujeres. Para ello, considera una muestra aleatoria de 38 hombres y 62 mujeres, observando que sólo a 15 hombres y 33 mujeres les había gustado su producto. Construir un intervalo de confianza al 99% de confianza para la diferencia de proporciones de hombres y mujeres a los que les gusta el producto. ¿Puede suponerse que el producto gusta por igual en hombres y mujeres?

Solución

```
> empresa <-
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/empresa.xls",
+    rownames=FALSE, header=TRUE, na="", sheet="Respuestas", stringsAsFactors=TRUE)
```

Dado que la hipótesis que se ha planteado se ha hecho la diferencia entre mujeres y hombres a los que les gusta el producto, es necesario hacer una recodificación de la variable Producto. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

```
> empresa <- within(empresa, {
+   Producto_rec <- Recode(Producto, '"Sí"="1Sí"; "No"="2No"', as.factor=TRUE)
+ })
```

A continuación realizamos un intervalo de confianza para la diferencia de dos proporciones. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para dos muestras

La pestaña Datos muestra dos listas con las variables cualitativas que incluye el conjunto de datos. De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Sexo) y de la segunda, la variable de interés (Producto_rec).

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado (99%).

```
> local({  .Table <- xtabs(~Sexo+Producto_rec, data=empresa)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.99, correct=FALSE)
+ })
```

```
Percentage table:
        Producto_rec
Sexo      1Sí  2No Total Count
  Hombre 39.5 60.5   100    38
  Mujer  53.2 46.8   100    62

    2-sample test for equality of proportions without continuity correction

data:  .Table
X-squared = 1.7851, df = 1, p-value = 0.1815
alternative hypothesis: two.sided
99 percent confidence interval:
 -0.3989753  0.1239329
sample estimates:
   prop 1    prop 2
0.3947368 0.5322581
```

Según los resultados, el intervalo de confianza para la diferencia de proporciones que buscamos es **(-0.3989753 0.1239329)**, el cual incluye al 0, por lo que se puede afirmar que el producto gusta por igual entre hombres y mujeres.

Ejercicio Guiado 4

En una experiencia genética se extraen 20 moscas de una caja experimental y se mide la longitud del ala de cada una. Se obtuvieron los siguientes valores:

93, 90, 97, 90, 93, 91, 96, 94, 91, 91, 88, 93, 95, 91, 89, 92, 87, 88, 90, 86

Suponiendo que la longitud del ala sigue una distribución Normal. Construir un intervalo de confianza al 99% de confianza para

a) La media μ

b) La varianza σ^2

Solución

```
> moscas <-
+   read.table("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/moscas.txt
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
```

a) Construir un intervalo de confianza al 99% de confianza para la media μ

Hay que obtener un intervalo de confianza cuando la varianza poblacional es desconocida

Accedemos al menú Test t para una muestra, seleccionando en el menú principal: Estadísticos/ Medias/ Test t para una muestra.

En este caso queremos obtener un intervalo de confianza, por lo que dejamos marcada la opción por defecto Media poblacional!=mu0 y como Hipótesis nula: mu=0.0. Lo único que tenemos que cambiar es el nivel de confianza (el valor se introduce en tanto por uno). Nivel de confianza: .99.

```
> with(moscas, (t.test(Longitud, alternative='two.sided', mu=0.0, conf.level=.99)))
```

```
    One Sample t-test

data:  Longitud
t = 139.01, df = 19, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 89.37195 93.12805
sample estimates:
mean of x
    91.25
```

El intervalo de confianza para la longitud media de las alas al 99% de confianza, es **(89.37195, 93.12805)**

b) Construir un intervalo de confianza para la varianza σ^2 al 99% de confianza.

Nota: R-Commander no incluye una función específica para el cálculo de intervalos de confianza en este tipo de situaciones. Por lo tanto calcularemos el intervalo de la siguiente forma

```
> n <- length(moscas$Longitud)
> varianza <- var(moscas$Longitud)
> L1 <- (n - 1) * varianza / qchisq(1-alpha / 2,n - 1)
> L2 <- (n - 1) * varianza / qchisq(alpha /2,n - 1)
> IC <- c(L1,L2)
> IC
```

```
[1]  4.244179 23.926166
```

El intervalo pedido es: **(4.244179, 23.926166)**

EJERCICIOS PROPUESTOS

Ejercicio Propuesto 1

En la tabla siguiente se muestran los salarios mensuales en euros de 10 trabajadores de Madrid y Barcelona.

Se pide:

a) Obtener un intervalo de confianza a un nivel del 89% para el salario medio entre ambas ciudades

b) Obtener un intervalo de confianza a un nivel del 90% para la diferencia media de salarios entre ambas ciudades. ¿Se pueden considerar iguales?

c) Obtener un intervalo de confianza a un nivel del 90% para la proporción de trabajadores en Barcelona.

Solución

```
> salario <-
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/salario.xlsx",
+    rownames=FALSE, header=TRUE, na="", sheet="Hoja1", stringsAsFactors=TRUE)
```

a) Obtener un intervalo de confianza a un nivel del 89% para el salario medio entre ambas ciudades

```
> with(salario, (t.test(Salario, alternative='two.sided', mu=0.0, conf.level=.89)))

	One Sample t-test

data:  Salario
t = 24.473, df = 9, p-value = 1.521e-09
alternative hypothesis: true mean is not equal to 0
89 percent confidence interval:
 1957.145 2262.855
sample estimates:
mean of x
     2110
```

Por lo que el intervalo de confianza es **(1957.145, 2262.855).**

b) Obtener un intervalo de confianza a un nivel del 90% para la diferencia media de salarios entre ambas ciudades. ¿Se pueden considerar iguales?

En primer lugar, comprobamos si las varianzas de ambas distribuciones son iguales.

```
> Tapply(Salario ~ Ciudad, var, na.action=na.omit, data=salario) # variances by group

Barcelona    Madrid
 95833.33  74666.67
```

```
> var.test(Salario ~ Ciudad, alternative='two.sided', conf.level=.95, data=salario)

	F test to compare two variances

data:  Salario by Ciudad
F = 1.2835, num df = 3, denom df = 5, p-value = 0.7512
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.1653207 19.1044044
sample estimates:
ratio of variances
          1.283482
```

El intervalo de confianza para el cociente de las varianzas es (0.2372666, 11.5686088). Dicho intervalo incluye al 1 entre sus posibles valores. Esto implica que a un nivel de confianza del 90% se puede suponer que el cociente entre las dos varianzas puede tomar el valor 1 o, lo que es lo mismo, que las dos varianzas son iguales.

Una vez se ha determinado la igualdad de las varianzas de ambas distribuciones, procedemos a calcular el intervalo de confianza para la diferencia de las medias propiamente dicho.

```
> t.test(Salario~Ciudad, alternative='two.sided', conf.level=.90, var.equal=FALSE,
+ data=salario)

	Welch Two Sample t-test

data:  Salario by Ciudad
```

```
t = -0.30574, df = 5.961, p-value = 0.7702
alternative hypothesis: true difference in means between group Barcelona and group Madrid is not equal
90 percent confidence interval:
 -429.5179  312.8513
sample estimates:
mean in group Barcelona    mean in group Madrid
               2075.000                2133.333
```

El intervalo de confianza a un 90% de confianza para la diferencia de las medias de salarios entre ambas ciudades es **(-403.3204, 286.6537)**. Como el 0 está dentro de este intervalo, tenemos suficiente evidencia muestral para decir que los salarios en ambas ciudades son iguales.

c) Obtener un intervalo de confianza a un nivel del 90% para la proporción de trabajadores en Barcelona.

En este caso nos encontramos ante un intervalo de confianza para la proporción.

Dado que la hipótesis que se ha planteado se ha hecho sobre Barcelona, no es necesario hacer ninguna recodificación de la variable.

```
> local({
+   .Table <- xtabs(~ Ciudad , data= salario )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.90, correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
Ciudad
Barcelona    Madrid
        4         6

    1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 0.4, df = 1, p-value = 0.5271
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.1942270 0.6483614
sample estimates:
  p
0.4
```

Por lo que el intervalo de confianza, a un nivel de confianza del 90% para la proporción de trabajadores en Barcelona es (0.1942270, 0.6483614).

Ejercicio Propuesto 2

Para comprobar si un determinado fertilizante puede mejorar la producción de manzanas, se selecciona una muestra aleatoria simple de 10 árboles. En la tabla siguiente se muestra el peso (en Kgr) de manzanas por árbol recogidas antes y después del tratamiento

Obtener un intervalo de confianza al 98% para la diferencia de los pesos medios producida antes y después del tratamiento.

Solución

```
> manzanas <-
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/manzanas.xlsx"
+    rownames=FALSE, header=TRUE, na="", sheet="Hoja1", stringsAsFactors=TRUE)
```

El intervalo de confianza al 98% para la diferencia de los pesos medios producida antes y después del tratamiento

```
> with(manzanas, (t.test(Antes, Después, alternative='two.sided', conf.level=.98, paired=TRUE)))
```

```
	Paired t-test

data:  Antes and Después
t = -2.1262, df = 9, p-value = 0.0624
alternative hypothesis: true mean difference is not equal to 0
98 percent confidence interval:
 -1.0006005  0.1406005
sample estimates:
mean difference 
          -0.43 
```

En la segunda parte de los resultados se incluye el intervalo de confianzaal 98% para la diferencia de las medias los pesos, que es (-1.0006005, 0.1406005). Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambos pesos son iguales.

Ejercicio Propuesto 3

En una muestra aleatoria de 150 personas con pelo oscuro se encontró que 90 de ellas tenían los ojos azules. Construir un intervalo de confianza al 95% para la proporción de individuos que teniendo pelo oscuro en la población posee ojos azules. ¿Son compatibles estos resultados con la suposición de que dicha proporción vale 2/3).

Solución

```
> propuesto <- 
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/propuesto3-1.x
+    rownames=FALSE, header=TRUE, na="", sheet="Hoja1", stringsAsFactors=TRUE)
```

Recodificar la variable

```
> propuesto <- within(propuesto, {
+   Azules_rec <- Recode(Azules, '"Sí"="1Sí"; "No"="2No"', as.factor=TRUE)
+ })
```

```
> local({
+   .Table <- xtabs(~ Azules_rec , data= propuesto )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.95, correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
Azules_rec
1Sí 2No 
 90  60 

	1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 6, df = 1, p-value = 0.01431
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5200492 0.6749568
```

```
sample estimates:
  p
0.6
```

El intervalo de confianza para la proporción de individuos que teniendo pelo oscuro en la población posee ojos azules, a un nivel del 95%, es (0.5200492, 0.6749568). Este resultado si es compatible con la suposición de que dicha proporción vale 2/3, ya que 2/3 pertenece al intervalo.

Ejercicio Propuesto 4

En una piscifactoría se desea comparar el porcentaje de peces adultos que miden menos de 20 cm con los que miden más de 40 cm. Para ello, se toma una muestra de 200 peces observando que 40 de ellos miden menos de 20 cm y una muestra de 200 peces de los que 57 miden más de 40 cm. Halla un intervalo de confianza para:

a) La diferencia de proporciones de peces adultos que miden más de 40 cm con los que miden menos de 20 cm al nivel de confianza del 0.95

b) La diferencia de proporciones de peces adultos que miden menos de 20 cm con los que miden más de 40 cm al nivel de confianza del 0.95

Solución

```
> propuesto4 <-
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/propuesto4-1.xl
+    rownames=FALSE, header=TRUE, na="", sheet="Hoja1", stringsAsFactors=TRUE)
```

a) La diferencia de proporciones de peces adultos que miden más de 40 cm con los que miden menos de 20 cm al nivel de confianza del 0.95

Recodificar la variable

```
> propuesto4 <- within(propuesto4, {
+   resultado_rec <- Recode(resultado, '"Sí"="1Sí"; "No"="2No"', as.factor=TRUE)
+ })
```

```
> local({  .Table <- xtabs(~medida+resultado_rec, data=propuesto4)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.95, correct=FALSE)
+ })
```

```
Percentage table:
           resultado_rec
medida       1Sí  2No Total Count
  masde40   28.5 71.5   100   200
  menosde20 20.0 80.0   100   200

    2-sample test for equality of proportions without continuity correction

data:  .Table
X-squared = 3.9332, df = 1, p-value = 0.04734
alternative hypothesis: two.sided
95 percent confidence interval:
 0.001410925 0.168589075
sample estimates:
prop 1 prop 2
 0.285  0.200
```

Intervalo de confianza para la diferencia de proporciones entre los peces adultos que miden más de 40 cm y los que miden menos de 20 cm al 95% (0.001410925, 0.168589075)

b) La diferencia de proporciones de peces adultos que miden menos de 20 cm con los que miden más de 40 cm al nivel de confianza del 0.95

Recodificar la variable

```
> propuesto4 <- within(propuesto4, {
+   media_rec <- Recode(medida, '"menosde20"="1menosde20"; "masde40"="2masde40";',
+   as.factor=TRUE)
+ })
```

```
> local({  .Table <- xtabs(~media_rec+resultado_rec, data=propuesto4)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.95, correct=FALSE)
+ })
```

```
Percentage table:
            resultado_rec
media_rec     1Sí  2No Total Count
  1menosde20 20.0 80.0   100   200
  2masde40   28.5 71.5   100   200

    2-sample test for equality of proportions without continuity correction

data:  .Table
X-squared = 3.9332, df = 1, p-value = 0.04734
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.168589075 -0.001410925
sample estimates:
prop 1 prop 2
 0.200  0.285
```

Intervalo de confianza para la diferencia de proporciones entre los peces adultos que miden menos de 20 cm y los que miden más de 40 cm al 95% (-0.168589075, -0.001410925).